

Geoparsing and Real-time Social Media Analytics

Technical and Social Challenges

Stuart E. Middleton

University of Southampton IT Innovation Centre
sem@it-innovation.soton.ac.uk @stuart_e_middle

ESRC Seminar: Microenterprise, technology and big data

11th October 2016

Overview

- Introduction
- Problem Statement
- Social Science Challenges
- Technical State of the Art
- Future Trends

Introduction

- IT Innovation Centre
- World-class application-driven R&D
 - Applied research and development with and for industry, commerce and the public sector
 - Collaborative research (supported by EC and UK programmes)
 - Client-funded research, development and consulting
- Research themes
 - Data science, big data and decision support
 - Human-centric computing and information sciences
 - Information security and risk management

Introduction

- Team of over 30 people
- Over the last five years
 - 42 major projects
 - 25 in the EC Framework Programme
 - over £2.25M of UK funding
 - over €10M of EC funding
 - working directly with tens of Universities
 - over 100 companies as partners and clients

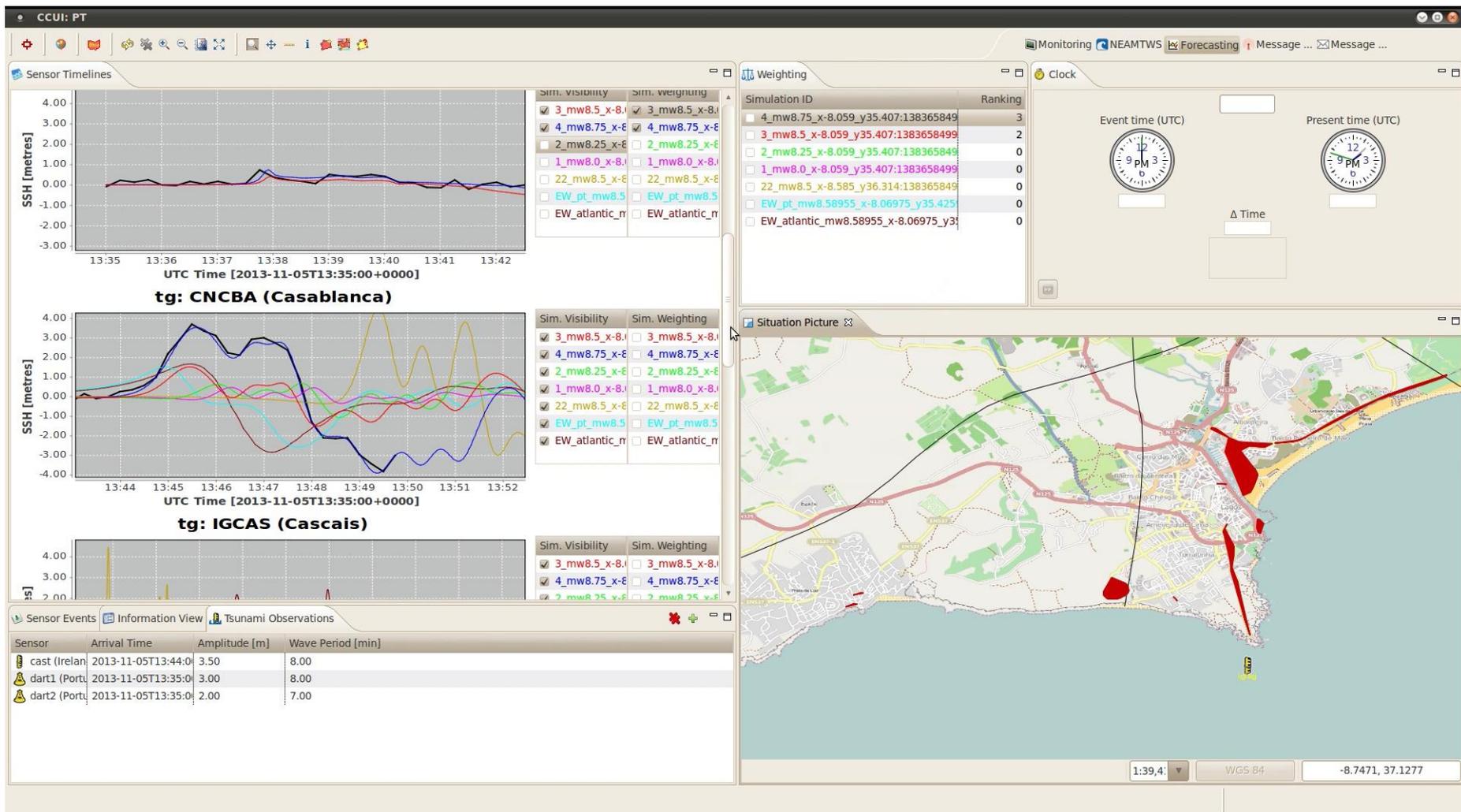
web: www.it-innovation.soton.ac.uk

twitter: [@IT_Innov](https://twitter.com/IT_Innov)

Problem Statement

- Social media analytics for crisis management
 - Real-time mixture of sensors and social media
 - Tide gauge sensors
 - Tsunami wave propagation simulations
 - You Tube videos of flooding
 - Extracting eyewitness reports from social media
 - Flooding damage, earthquake shakes ...
 - Spatial analysis of eyewitness media
 - Damage assessments, confirmation of events, cries for help
 - Decision support for civil protection agencies
- TRIDEC project www.tridec-online.eu
 - Tsunami Early Warning

Problem Statement



Cite: Middleton, S.E. Sabeur, Z.A., Löwe, P., Hammitzsch, M., Tavakoli, S., Poslad, S. "Multi-disciplinary Approaches To Intelligently Sharing Large-volumes Of Real-time Sensor Data During Natural Disasters", Data Science Journal, Volume 12, pages 109-113 (2013)

Problem Statement

- Social media analytics for journalistic verification
 - Breaking news decision timeframes
 - 10 minute window to find eyewitness media and verify it
 - After 20+ minutes its been broadcast and is not new anymore
 - Real-time eyewitness media
 - Twitter, Facebook, Instagram, You Tube, Periscope
 - User generated content (UGC) often at the scene of an event
 - Eyewitness images & videos, incidents & damage reports ...
 - Supporting a manual verification process
 - Journalist best practice
- REVEAL project www.revealproject.eu
 - User generated content verification

Problem Statement



(a)



(b)



(c)

(a) Digitally tampered original - photograph of IAF F-16 deploying a single flare over Southern Lebanon. The flare was digitally duplicated to make it appear that several missiles were being fired.

(b) Historical image recycled - picture of a storm descending in New York, purported to be captured during Hurricane Sandy, while it was featured in Wall Street Journal one year earlier.

(c) Completely fake - Image claiming to depict the solar eclipse in 2012, while it was in fact digital artwork originally published on DevianArt.

Cite: C. Boididou et al., "Challenges of computational verification in social multimedia"
In Proceedings of the 23rd International Conference on World Wide Web (2014)

Social Science Challenges

- The rise of the general public as live reporters
 - Smartphone + social media = UGC
 - User Generated Content (UGC) is fast and free
 - Transforming the way journalists work
 - Journalists are telling stories with UGC
 - Why send a journalist to the scene? Is UGC enough?
 - Faster (breaking news) and more local (niche content)
 - Editorial verification costs are rising though
 - Unverified content = more checking = more staff time

Social Science Challenges

- The rise of the general public as live reporters
 - Micro-enterprise context
 - Low cost business feedback easy via social media
 - Content costly (effort) to verify but automation helps
 - Risks
 - Verification costs to avoid propagating rumours
 - Analysis staff (over) exposure to unfiltered UGC
 - *Graphic content (sexual, violent, disturbing)*
 - *Worker mental health*

Social Science Challenges

- Collaborative news desk remotely distributed
 - Modern news desks telework via smartphones
 - Verification of UGC is a collaborative effort
 - Contact the source
 - Do background checks
 - Look at contextual material
 - Bring in domain expertise
 - Collaboration is key to decision making
 - Colleagues each provide a different viewpoint
 - *Local knowledge (places, people, organisations)*
 - *Cultural norms, writing styles, common vocabulary*
 - Each viewpoint contributes evidence for the overall verification decision

Social Science Challenges

- Collaborative news desk remotely distributed
 - Micro-enterprise context
 - Social media and smartphones can facilitate distributed networks of contacts / workers
 - Reduced travel costs
 - Homeworking
 - Allows collaborations previously too expensive to facilitate
 - Risks
 - 24/7 culture
 - Worker stress

Social Science Challenges

- Evidence-based approach to decision support
 - Journalists need evidence to make decisions
 - Black boxes are not trusted
 - Each journalist's reputation is on the line
 - Automate the mundane stuff
 - Empower staff to focus on the difficult stuff
 - Algorithms aggregate, cluster and summarize easily
 - People add value on the hard subjective decisions
 - Show evidence behind each algorithms decision
 - For example show the journalist the social media posts in clusters that map to a final evidential conclusion

Social Science Challenges

- Evidence-based approach to decision support
 - Micro-enterprise context
 - Social media analytics can inform & focus advertising / user education / marketplace strategy
 - *App and web-site click through data, user rating feedback*
 - *Social media sentiment analysis during a product launch*
 - Risks
 - Algorithmic results get manipulated
 - *Bot farms (e.g. US Politics) and governments (e.g. Russian/Ukraine) are spamming biased content*
 - *Fake social media posts (e.g. likes, comments) and ratings (e.g. user reviews)*
 - *Easy for humans to spot fakes but algorithms will blindly aggregate this data*
 - *A few trusted reports > millions of 'echo chamber' posts*

Social Science Challenges

- Protecting users right to privacy
 - Our lives are shared online
 - Evidence available from status/timeline, friends, location, interests/groups
 - Opt out?
 - Your friends activity can still be used to find you ...
 - Right to be forgotten?
 - Impossible to forget if your data has already been aggregated ...

Social Science Challenges

- Protecting users right to privacy
 - Micro-enterprise context
 - Being seen to respect user privacy is customer friendly
 - Risks
 - Losing user confidence on privacy issues
 - Customers deleting accounts or simply stopping using them

Technical State of the Art

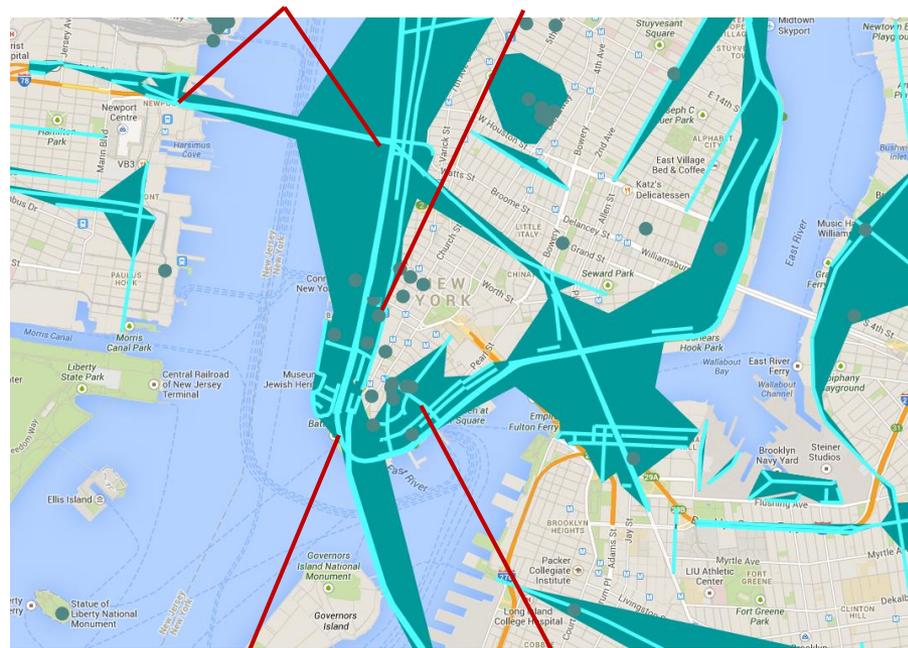
• Geoparsing

Holland Tunnel

Ground Zero

Holland Tunnel

Ground Zero



Battery Place

Wall Street

Battery Place

Wall Street

Key  Expert post-event assessment storm surge inundation area

 Place flooded tweet(s)

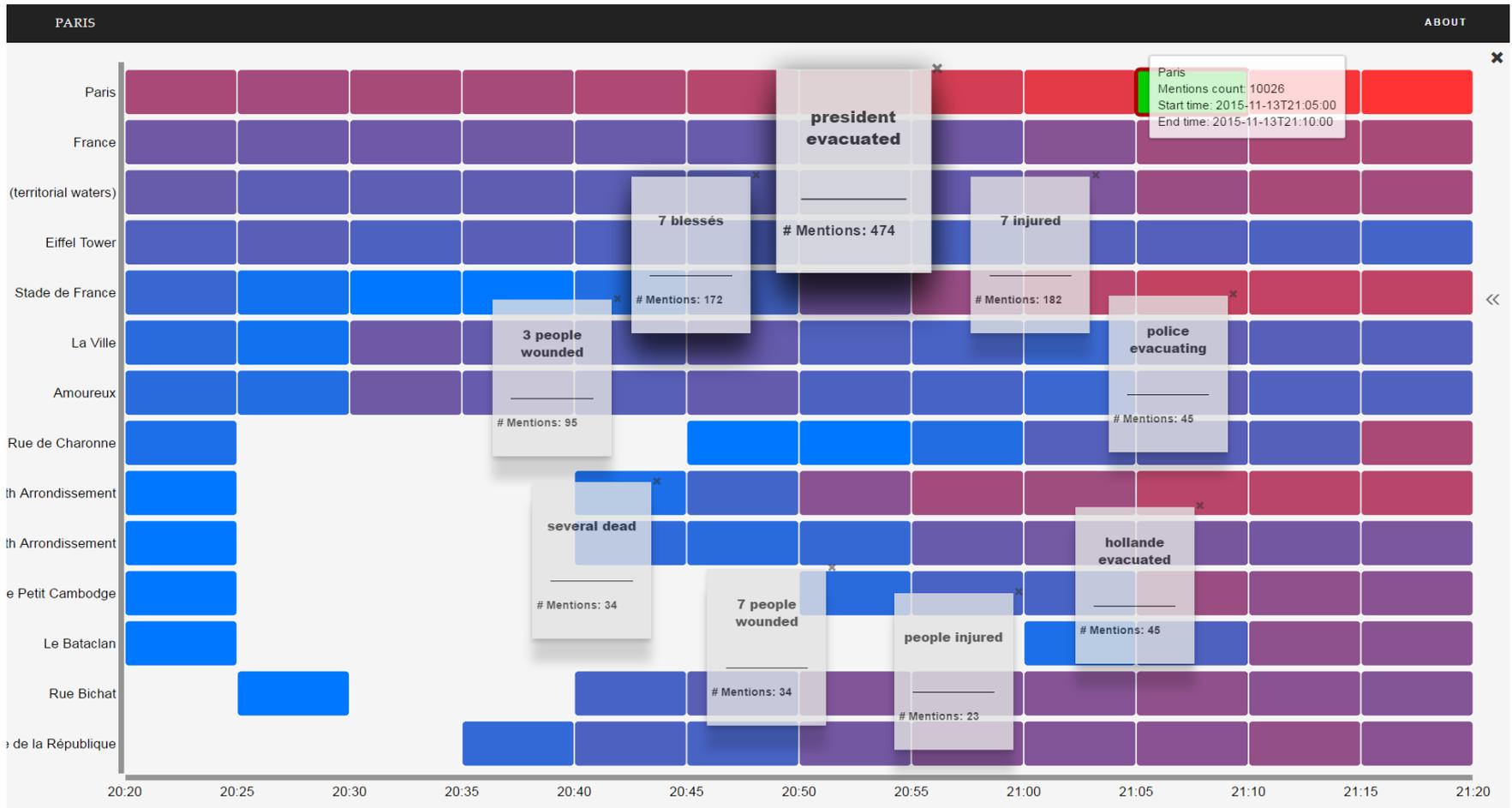
 Clustered flood reports

 Street flooded tweet(s)

Cite: Middleton, S.E. Middleton, L. Modafferi, S. "Real-time Crisis Mapping of Natural Disasters using Social Media", Intelligent Systems, IEEE , vol.29, no.2, pp.9,17, Mar.-Apr. 2014

Technical State of the Art

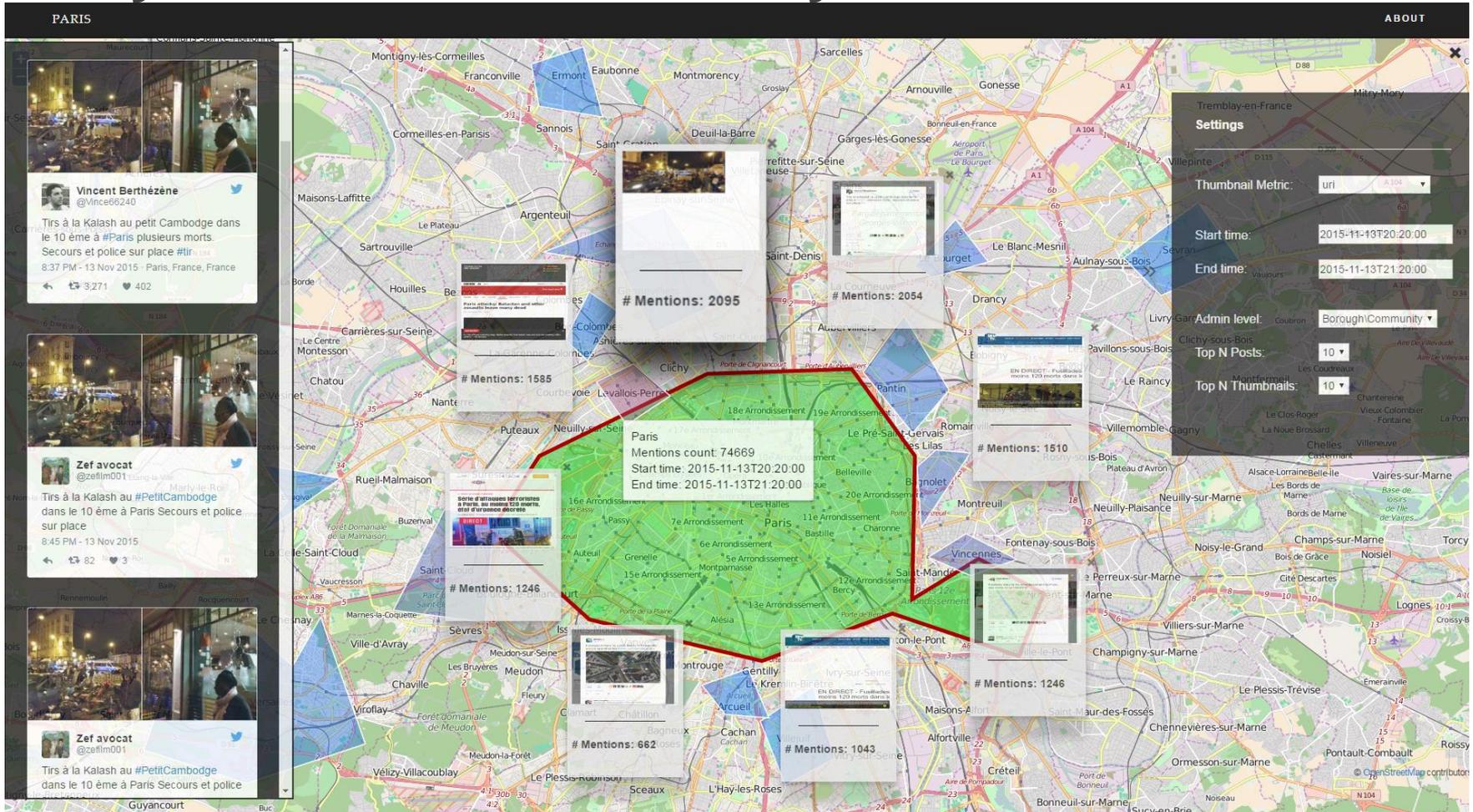
- Fact extraction



Cite: Middleton, S.E. "Extracting Attributed Verification and Debunking Reports from Social Media: MediaEval-2015 Trust and Credibility Analysis of Image and Video", MediaEval-2015, Wurzen, Germany (2015)

Technical State of the Art

- Eyewitness & Trust analysis



Cite: Middleton, Middleton, S.E. Krivcovs, V. "Geoparsing and Geosemantics for Social Media: Spatio-Temporal Grounding of Content Propagating Rumours to support Trust and Veracity Analysis during Breaking News", ACM Transactions on Information Systems (TOIS), 34, 3, Article 16 (2016)

Future Trends

- Social media
 - Big datasets controlled by a few companies
 - Unregulated big data mining
 - Disambiguation & de-anonymization of our profiles across social media sites will become routine
- Internet of Things
 - Mobile phones, smart buildings, smart environments
 - Sensors, cameras, GPS, audio, CCTV
 - Data fingerprinting all our daily lives

Future Trends

- Human / bot collectives
 - Crowd-sourcing contribution, editing and verification behaviours
 - Bots and humans working together
- Cyber-crime
 - Massive data breaches
 - Passwords/logins and personal activity/behaviour traces
 - Big datasets are a target for cybercriminals
 - Dark web marketplaces
 - Identify theft & ransomware at an industrial scale



Thanks you for your attention!

Any questions?

Try our Journalist Decision Support System (JDSS) for free!

https://reveal-jdss.it-innovation.soton.ac.uk/reveal_journalists_dss

Dr Stuart E. Middleton

University of Southampton IT Innovation Centre

email: sem@it-innovation.soton.ac.uk

web: www.it-innovation.soton.ac.uk

twitter: [@stuart_e_middle](https://twitter.com/stuart_e_middle) [@IT_Innov](https://twitter.com/IT_Innov)